

RESEARCH

Open Access



Genome-wide association study of never-smoking non-drinking young adults developing oral squamous cell carcinoma

Guillaume B. Cardin^{1,2}, Alicia Pellerin-Viger^{1,2}, Monique Bernard^{1,2}, Fiona Kessai^{1,2}, Tareck Ayad^{1,3}, Eric Bissada^{1,3}, Louis Guertin³, Paul Tabet^{1,3}, Olguta-Ecaterina Gologan⁴, Edith Filion⁵, Phuc Felix Nguyen-Tan⁵, Antoine Desilets^{1,2,6}, Alexandre Pellan Cheng^{1,2,7}, Francis Rodier^{1,2,8}, Houda Bahig^{1,2,5} and Apostolos Christopoulos^{1,2,3*}

Abstract

Background Germline risk variants for oral squamous cell carcinoma (OSCC) in never-smoking non-drinking (NSND) young adults (YA) are poorly characterized and genome-wide association studies focusing on NSND YA OSCC are lacking. This study aims to evaluate if rare germline variants are associated with NSND YA OSCC.

Methods We conducted a retrospective, single-center Canadian cohort study of patients who underwent primary surgery for OSCC between 2010 and 2024. Genome-wide association analyses were performed for three comparisons: 1- NSND YA ($n = 10$) vs. other OSCC ($n = 150$), 2- all-ages NSND ($n = 29$) vs. ever-smoker and/or drinker ($n = 131$), and 3- any risk factor YA OSCC ($n = 22$) vs. age ≥ 45 ($n = 138$). A polygenic risk score for age at OSCC diagnosis was derived from clumping + thresholding.

Results We included 160 patients of European ancestry with a diagnosis of OSCC: 22 YA and 29 NSND patients; 10 patients were both YA and NSND. YA cases predominantly involved the oral tongue. One rare variant, rs191595756, showed a suggestive association with NSND YA OSCC ($p = 3.3e-06$), but did not reach genome-wide significance. This variant is located in an intron of LINC00944 and in an exon of LINC02824. No rare variant signals were supported in the all-ages NSND and any risk factor YA OSCC. None of the identified variants is reported in ClinVar. An exploratory polygenic risk score composed of 13 index variants was constructed. It is independent of NSND or ever-smoker and/or drinker status. It is also independent of oral tongue or other oral cavity primary cancer site.

Conclusions This is the first study investigating genome-wide associations in NSND YA OSCC. No single germline variant of genome-wide significance associated with NSND YA OSCC occurrence was found, suggesting a polygenic architecture. Larger, multicentric studies are needed to validate a polygenic risk score specifically predictive of NSND YA OSCC risk.

Keywords Genome-wide association study, Never-smokers, Oral Cancer, Polygenic risk score, Young adults

*Correspondence:
Apostolos Christopoulos
a.christopoulos@umontreal.ca

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

An increase in the incidence of oral squamous cell carcinoma (OSCC) in young adults (YA), with a predominance of oral tongue cancers, has been reported since the 1980s and has since been confirmed in multiple reports worldwide over recent decades [1–9]. The age threshold defining YA OSCC varies across studies and is most commonly set at 40 or 45 years [10]. While tobacco and alcohol use are frequent in older patients with OSCC, their prevalence and contribution are lower in YA OSCC [10]. This has led to the recognition of a YA never-smoking non-drinking (NSND) OSCC subtype, although underlying causal mechanisms remain elusive.

As young patients are involved, the role of viruses has been extensively studied. Oncogenic viruses are common etiologic factors in several head and neck cancer subsites, but previous sequencing-based efforts to identify viral drivers in YA OSCC tumors have not been unsuccessful [11]. Human papillomavirus (HPV) can sometimes be detected in YA OSCC, but appears unlikely to have an important etiologic role [12–14]. Studies on oral tongue and OSCC looking at HPV directly, and not p16(INK4a) as a surrogate, indicate that its incidence is rare in this disease. A cohort of 32 patients, with 6 cases under 30 years old, showed a frequency of 0/32 of high-risk HPV [15]. Another study looked at 78 oral tongue cancer patients detected HPV in only one case, but in none of the 11 patients under 40 years old [16]. In a cohort of 95 NSND OSCC patients, Epstein-Barr virus (EBV) and Merkel cell polyomavirus (MCV) were also evaluated with EBV only detected in a single case, and with no cases of MCV [17].

A family history of cancer in young relatives has been associated with both YA OSCC and head and neck cancer more generally [10, 18], suggesting that genomic susceptibility could play a role in YA OSCC. From a genetic standpoint, tumor exome analyses of non-smoking OSCC and YA non-smoking OSCC have been published [19, 20], but very little is known about potential germline at-risk variants in those patients. Previous genome-wide association studies have identified OSCC susceptibility loci located on multiple chromosomes [21–24]. However, these analyses largely reflect tobacco- and alcohol-related carcinogenesis and reports testing for age-associated at-risk variants are missing. To date, no polygenic risk score (PRS) predicting age at OSCC diagnosis has been reported.

Evaluating germline at-risk variants associated with YA OSCC – particularly in NSND individuals – remains an important gap in current knowledge. The main objective of this exploratory study is to evaluate, using whole genome sequencing (WGS), whether rare germline variants are associated with NSND YA OSCC. A secondary objective of this study was to construct a PRS predicting

age at OSCC diagnosis, to evaluate whether genetic susceptibility patterns differ between NSND and to YA OSCC and ever-smoker and/or drinker (ESD) OSCC. While tumor genomics have been partially explored, germline susceptibility in NSND YA OSCC remains almost completely uncharacterized, justifying an exploratory genome-wide association study on a small study cohort.

Methods

Study design and participants

A total of 160 OSCC patients of European ancestry were included in the first four years of the Marathon of Hope Cancer Centres Network (MOHCCN), a pan-cancer initiative aiming to promote access to WGS and build a clinico-genomic database encompassing 15,000 Canadians with a diagnosis of cancer [25]. The OSCC cohort included patients 18 years or older who underwent surgery at the Centre Hospitalier de l'Université de Montréal (CHUM) between 2010 and 2024, and provided written consent to participate in the institutional head and neck cancer biobank at the CHUM Research Center (CRCHUM). This study and the biobank were approved by CRCHUM's institutional research ethics board (#21.137 and #09.254).

Demographics and clinical parameters

Demographics and clinical parameters were obtained from clinical records, including sex, age at diagnosis, tobacco use, alcohol use, primary tumor site (oral tongue, floor of mouth, mandibular alveolus, retromolar trigone, buccal mucosa, mucosal lip, hard palate, maxillary alveolus), self-identified genetic ancestry group, and pathologic stage. For tobacco use, patients were categorized as smokers if they were current or former smokers at the time of surgery, with the exception of participants with less than 1 pack-year history. For alcohol use, patients with International Classification of Diseases, Tenth Revision (ICD-10) Z72.1 (alcohol use) or F10.2 (alcohol dependence) were classified as “drinkers”. Patients under 45 years of age at OSCC diagnosis were classified as YA. Pathologic stage was recorded according to the 8th edition of the American Joint Committee on Cancer (AJCC). Patients from non-European ancestry were excluded to prevent introduction of false positives due to linkage disequilibrium. The majority of included patients were French Canadians.

Germline DNA extraction and sequencing

Peripheral blood samples were obtained at the time of surgery, and peripheral blood mononuclear cells were separated using Ficoll and stored for future analyses. Germline DNA was extracted from peripheral blood mononuclear cells using the PureLink Genomic DNA

Mini Kit (Thermo Fisher Scientific). For one participant, germline DNA was extracted from a negative control lymph node using the AllPrep DNA/RNA FFPE Kit (Qia-gen). Nucleic acid quality was assessed using the 4150 TapeStation System (Agilent) prior to library preparation. WGS was performed on germline DNA, following Genome Analysis Toolkit's (GATK) Best Practices and procedures, with a target of 30X deduplicated coverage per sample, using an Illumina NovaSeq 6000 instrument.

Genotyping

Quality-trimmed and adapter-clipped reads were aligned to the GCA_000001405.15 GRCh38_no_alt_analysis_set.fna human genome assembly using BWA-MEM [26]. Germline variant calls were generated combining Strelka2, VarScan 2, and VarDict calls in one Variant Call Format (VCF) file per participant [27–29]. Single sample VCF files were merged, indexed and filtered to exclude variants with read depth below 25 or genotype quality below 20 using bcftools [30]. The resulting VCF file was converted to a BED/BIM/FAM fileset, including only variants from chromosome 1 to 22 and filtering out multiallelic variants using PLINK version 2.0 [31]. Variants on the X and Y chromosomes were removed as a best practice to prevent confounding by sex. Variants with complete cohort minor allele frequency (MAF) < 0.009 (fewer than 3 occurrences in the full cohort) were also excluded, as they have low statistical power and high false positive rate.

Genome-wide association analysis of NSND YA, all-ages NSND and any risk factor YA OSCC

Association analyses were conducted separately for the NSND YA group, the all-ages NSND group and the YA any risk factor group, using other participants of the cohort as controls and applying Fisher's exact test in PLINK version 1.9 [32]. Index variants were identified using PLINK's --clump command with --clump-r2 0.1 and --clump-kb 250. Participant genotypes for variants with $p < 1.0\text{e-}05$ were extracted using PLINK and confirmed, as a quality control step, on the BAM files using IGV version 2.16.2 [33]. Manhattan plots were constructed using the qqman package in R. The All of Us data browser was used on October 21, 2025, to identify the European and the highest reported MAF from every genetic ancestry group of the All of Us cohort [34]. Rare variants were defined as having a highest reported MAF, from every genetic ancestry group of the All of Us cohort, under 0.01. The sample size ($n = 10$) is very small for the genome-wide association analysis of the NSND YA OSCC group, and the analysis is exploratory for this rare disease group, with a suggestive significance threshold of $p < 1.0\text{e-}05$.

Genome-wide association analysis and PRS for age at OSCC diagnosis

Association analyses with numerical age at diagnosis (in years, rounded to two decimals) was done on the full cohort using Wald test in PLINK version 1.9. Index variants identification, and BAM validation were realized, as described for association analysis. PRS was calculated by clumping + thresholding (C + T) using PLINK's --score command with the β -value of the regression. PLINK's --q-score-range command testing addition of variants with $p < 1.0\text{e-}05$, one by one, starting with the lowest p -value, allowed identification of the threshold maximizing R^2 [35]. PRS was transformed as "log -(PRS-1)" to test whether the slopes and intercepts of the regression between OSCC subgroups were significantly different, using Prism version 10.2.3. For the overall survival analysis, a multivariate Cox proportional hazards analysis was used to evaluate hazard ratios using JMP. Comparison of PRS distributions for TNM status was evaluated using Welch's ANOVA test. Since a validation cohort was unavailable, PRS validation was done using permutation analysis, randomly permuting phenotypes 10,000 times, with the empirical p -value calculated using R.

Results

Study cohort

A total of 160 patients with a diagnosis of OSCC were included at CHUM and included in the MOHCCN Canadian pan-cancer cohort, with clinical characteristics detailed in Table S1. All ($n=160$; 100%) were of self-identified European ancestry and 22 (14%) were YA (Table 1). Among those YA patients, 19 (86%) were male, 10 (45%) were NSND, and 18 (82%) had an oral tongue primary cancer. For the 138 patients aged ≥ 45 years, 92 (67%) were male, 19 (14%) were NSND, and 61 (44%) had an oral tongue primary cancer. Cancer stage distribution was slightly more aggressive in the ≥ 45 years old group.

Genome-wide association analysis of YA and/or NSND OSCC

The first objective of this study was to evaluate whether rare germline variants are associated with the occurrence of NSND YA OSCC. A genome-wide association analysis comparing NSND YA OSCC patients ($n=10$) to other OSCC cases ($n=150$), all-ages NSND OSCC patients ($n=29$) to ESD OSCC cases ($n=131$) and YA OSCC patients ($n=22$) to other OSCC cases ($n=138$) was performed. These analyses revealed no genome-wide significant hits, and only one rare variant with suggestive significance in the NSND YA analysis.

A total of 655 730 (NSND YA), 646 052 (NSND) and 648,351 (YA) index variants were identified after clumping. Manhattan plots for each differential analysis are shown in Fig. 1, with a red line showing the genome-wide

Table 1 Clinical and demographic characteristics of young adults (YA) and older oral squamous cell carcinoma (OSCC) patients in the study cohort

	YA OSCC < 45 years old (n = 22) No. (%)	Older OSCC ≥ 45 years old (n = 138) No. (%)
Sex		
Female	3 (14)	46 (33)
Male	19 (86)	92 (67)
Risk factors ^a		
None	10 (45)	19 (14)
Tobacco use only	12 (55)	63 (46)
Tobacco and alcohol use	0 (0)	53 (38)
Alcohol use only	0 (0)	3 (2)
Primary site		
Oral tongue	18 (82)	61 (44)
Floor of mouth	0 (0)	36 (26)
Mandibular alveolus	1 (5)	17 (12)
Retromolar trigone	3 (14)	11 (8)
Buccal mucosa	0 (0)	9 (7)
Mucosal lip	0 (0)	2 (1)
Hard palate	0 (0)	1 (1)
Maxillary alveolus	0 (0)	1 (1)
Cancer stage		
I	3 (14)	11 (8)
II	7 (32)	29 (21)
III	3 (14)	31 (22)
IVa/IVb	9 (41)	67 (49)
Genetic ancestry group		
European	22 (100)	138 (100)

^a Positive tobacco use is defined as a history of smoking (current or former) with ≥ 1 pack-year. Positive alcohol use is identified by ICD-10 diagnostic codes Z72.1 (alcohol use) or F10.2 (alcohol dependence)

significance threshold (5×10^{-8}) and a blue line showing the suggestive significance threshold (1×10^{-5}). This suggestive significance threshold was used to identify potential rare germline at-risk variants which could be confirmed in larger validation cohorts. Genomic inflation was also evaluated using Q-Q plots for each analysis (Figure S1).

The next step was to assess whether the identified variants were rare variants. To do so, we validated the highest reported MAF from every genetic ancestry group of the All of Us cohort, using a threshold of $MAF < 0.01$ (Table 2). Identified variants with $p < 1.0 \times 10^{-5}$ for NSND YA OSCC (Table S2) are rs149462127 (chr5:166940729-C-T, intergenic region, nearest gene: ENSG00000298538), rs16946701 (chr15:63255812-G-A, RAB8B intron), rs191595756 (chr12:126719941-A-G, LINC02824 exon and LINC00944 intron), rs113104790 (chr16:74142466-C-T, PSMD7-DT intron), and rs5879963 (chr6:129703003-CA-C, ARHGAP18 intron). Only rs191595756 can be considered a rare variant with the defined threshold. Its genotype distribution in

the NSND YA OSCC group is 6 homozygous reference alleles patients, 3 heterozygous patients and 1 homozygous alternate allele patients ($MAF = 0.25$), while for other OSCC cases it is 149 homozygous reference alleles patients, 1 heterozygous patients and no homozygous alternate allele patients ($MAF = 0.0033$). The control group distribution is consistent with the European MAF of 0.0035 reported in the All of Us cohort. For rs16946701 and rs113104790, the MAF in the other OSCC group was much lower than the reported European MAF in the All of Us cohort, by factors of 4.5 and 4.2 respectively, suggesting that those variants are false positives. Genomic deflation observed in the Q-Q plot is expected with the small sample size (Figure S1A). None of the identified variants is reported in ClinVar.

Variants with $p < 1.0 \times 10^{-5}$ for all-ages NSND OSCC (Table S3) are rs10495194 (chr1:222560629-C-T, TAF1A intron), rs4419048 (chr16:88559144-C-T, ENSG00000289151 intron), rs60356568 (chr2:232236404-G-A, DIS3L2 intron), rs11772757 (chr7:132378734-T-G, PLXNA4 intron), rs77235807 (chr3:29659276-G-C, RBMS3 intron and ENSG00000283563 intron), and rs12450836 (chr17:65196941-G-A, RGS9 intron). The absence of tail in the Q-Q plot suggests that there are no strong genetic association with NSND OSCC in the study cohort (Figure S1B). These 6 variants are not rare variants and their MAF in the ESD group were consistent with the European MAF reported in the All of Us cohort. None of the identified variants is reported in ClinVar.

Variants with $p < 1.0 \times 10^{-5}$ for YA OSCC (Table S4) are rs113765368 (chr9:122461042-A-G, ENSG00000234156 intron), rs17142701 (chr6:6936428-A-G, intergenic region, nearest gene: RN7SL554P), rs16946701 (chr15:63255812-G-A, RAB8B intron), rs56353403 (chr11:12079909-TG-T, ENSG00000305657 exon), and rs79853253 (chr21:15290364-G-A, ENSG00000288578 intron and ENSG00000300143 intron). These 5 variants are not rare variants. For rs113765368, rs17142701 and rs113104790, the ≥ 45 years old OSCC group MAF was much lower than the reported European MAF in the All of Us cohort, by factors of 5.1, 6.1 and 8.3 respectively, suggesting that those variants are false positives. The absence of tail in the Q-Q plot suggests that there are no strong genetic association with YA OSCC in the study cohort (Figure S1C). None of the identified variants is reported in ClinVar.

PRS for age at OSCC diagnosis

To eliminate the bias of the 45-years-old threshold of the YA group, we chose to perform a genome-wide association analysis with numerical age at diagnosis. We then compared the correlation between PRS and age at diagnosis for the NSND and the ESD subgroups, finding no

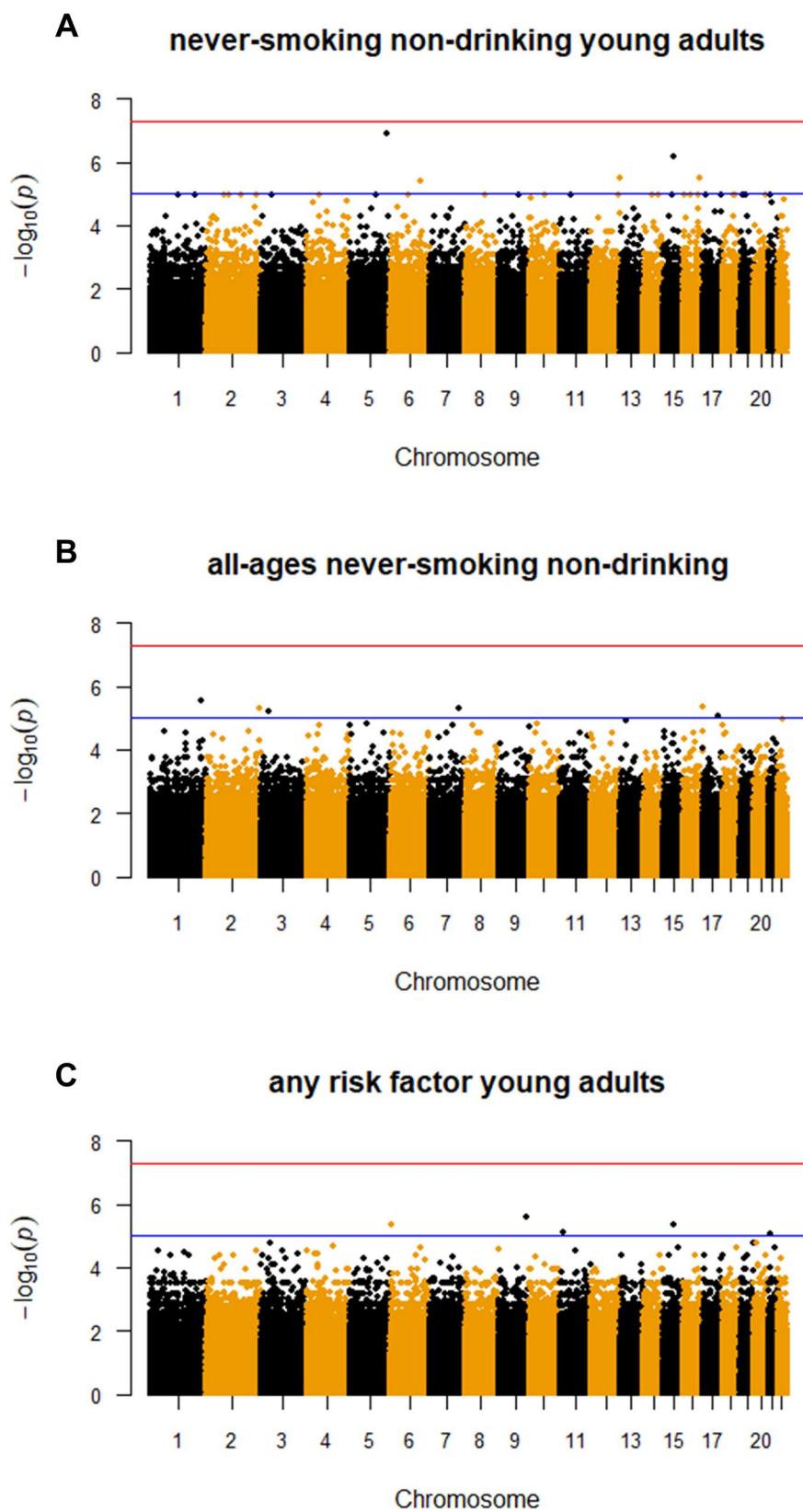


Fig. 1 Manhattan plots of the genome-wide association study. Manhattan plots for never-smoking non-drinking (NSND) young adults (YA) oral squamous cell carcinoma (OSCC) (**A**), All-ages NSND OSCC (**B**), and Any risk factor YA OSCC (**C**). Each dot represents a single index variant. The red line indicates the genome-wide significance threshold (5.0×10^{-8}). The blue line indicates the suggestive significance threshold (1.0×10^{-5})

Table 2 Germline variants with $p < 1e-05$ identified in the genome-wide association study comparing never-smoking non-drinking (NSND) and/or young adults (YA) oral squamous cell carcinoma (OSCC) patients with other OSCC patients

Test group	Germline variant ^a	OSCC cohort			All of Us reference cohort highest MAF ^d
		Test group MAF ^c	Others MAF	p	
NSND YA OSCC	rs149462127	0.35	0.010	1.3e-07	0.011
NSND YA OSCC	rs16946701	0.30	0.0067	7.1e-07	0.20
NSND YA OSCC	rs191595756	0.25	0.0033	3.3e-06	0.0035
NSND YA OSCC	rs113104790	0.25	0.0033	3.3e-06	0.015
NSND YA OSCC	rs5879963 ^b	0.55	0.10	3.9e-06	0.31
All-ages NSND OSCC	rs10495194	0	0.23	2.7e-06	0.18
All-ages NSND OSCC	rs4419048 ^b	0.26	0.046	4.4e-06	0.29
All-ages NSND OSCC	rs60356568	0.31	0.073	4.9e-06	0.50
All-ages NSND OSCC	rs11772757	0.017	0.26	5.0e-06	0.22
All-ages NSND OSCC	rs77235807	0.33	0.08	6.2e-06	0.20
All-ages NSND OSCC	rs12450836	0.26	0.050	8.3e-06	0.40
YA OSCC	rs113765368	0.18	0.0072	2.6e-06	0.11
YA OSCC	rs17142701	0.16	0.0036	4.3e-06	0.058
YA OSCC	rs16946701	0.16	0.0036	4.3e-06	0.20
YA OSCC	rs56353403 ^b	0.30	0.66	7.8e-06	0.63
YA OSCC	rs79853253	0.27	0.043	8.6e-06	0.17

^a Every suggestively significant germline variant with $p < 1e-05$ is shown^b Minor allele is the GRCh38 reference allele^c Minor allele frequency^d Highest reported MAF from every genetic ancestry group of the All of Us cohort

significant differences. The results of this analysis are exploratory, as the cohort is too small to draw meaningful conclusions. Larger, multicentric studies are needed to validate a polygenic risk score specifically predictive of NSND YA OSCC risk.

A total of 630 739 index variants were identified by clumping. The PRS for age at OSCC diagnosis was calculated. The calculation of R^2 following sequential addition of variants with $p < 1.0e-05$ in the model, starting with the lowest p -value, allowed the identification of the threshold that maximized R^2 . Using permutation analysis,

Table 3 Germline variants included in the polygenic score for age at oral squamous cell carcinoma (OSCC) diagnosis

Germline variant ^a	OSCC cohort			All of Us reference cohort highest MAF ^d
	MAF ^c	R^2	p	
rs17142701	0.025	0.169	6.6e-08	0.058
rs16946701	0.025	0.154	3.0e-07	0.20
rs199603591	0.072	0.136	1.7e-06	0.051
rs80326268	0.041	0.133	2.1e-06	0.039
rs34440545	0.39	0.132	2.4e-06	0.48
rs79949466	0.084	0.131	2.6e-06	0.16
rs116937215	0.019	0.130	2.9e-06	0.018
rs74590769	0.038	0.125	4.7e-06	0.041
rs75127284 ^b	0.069	0.124	4.9e-06	0.28
rs116985485	0.019	0.124	4.9e-06	0.080
rs568878631	0.028	0.124	5.1e-06	0.0076
rs12625513	0.041	0.124	5.1e-06	0.15
rs34497370	0.031	0.123	5.4e-06	0.033

^a Every suggestively significant germline variant with $p < 1e-05$ is shown^b Minor allele is the GRCh38 reference allele^c Minor allele frequency^d Highest reported MAF from every genetic ancestry group of the All of Us cohort

no permuted test statistics exceeded the observed PRS association using 10,000 permutations (empirical p -value = 9.999e-05). The resulting PRS included the following 13 index variants (Table 3 and Table S5): rs17142701 (chr6:6936428-A-G, intergenic region, nearest gene: RN7SL554P), rs16946701 (chr15:63255812-G-A, RAB8B intron), rs199603591 (chr20:39687424-TTC-T, intergenic region, nearest gene: ENSG00000306887), rs80326268 (chr16:54488136-C-T, intergenic region, nearest gene: ENSG00000286845), rs34440545 (chr7:49593616-CTAT-C, intergenic region, nearest gene: ENSG00000286995), rs79949466 (chr5:37429113-A-G, WDR70, intron), rs116937215 (chr9:13592714-T-A, ENSG00000226197 intron), rs74590769 (chr6:151377741-A-G, ZBTB2 intron), rs75127284 (chr18:34506592-A-C, DTNA intron), rs116985485 (chr7:101536460-G-A, COL26A1 intron), rs568878631 (chr1:67730331-C-T, GNG12 intron), rs12625513 (chr20:62844309-G-A, TCFL5 intron), rs34497370 (chr10:61377055-G-A, intergenic region, nearest gene: ENSG00000305028). From those, rs568878631 is a rare variant. Its genotype distribution in the OSCC cohort is 151 homozygous reference alleles patients, 9 heterozygous patients and no homozygous alternate allele patients (MAF = 0.028). None of the 13 index variants is reported in ClinVar.

The regression between age at diagnosis and PRS was evaluated for NSND and ESD OSCC patients separately. No significant difference was observed between NSND and ESD OSCC for the slope ($p = 0.12$) and intercept

($p = 0.16$) of the association between PRS and age at diagnosis (Figure S2). As the distribution of PRS was skewed, association between $-\log(-\text{PRS}-1)$ and age at diagnosis was also tested (Fig. 2A). No significant difference was observed between both slopes ($p = 0.996$) and intercepts ($p = 0.83$). The regression between age at diagnosis and $-\log(-\text{PRS}-1)$ was evaluated for oral tongue and other oral cavity squamous cell carcinoma patients (Fig. 2B). No significant difference was observed between both slopes ($p = 0.96$) and intercepts ($p = 0.20$). There was also no association between $-\log(-\text{PRS}-1)$ and overall survival (Table S6) and pathological TNM (Figure S3).

Discussion

In this single cohort study, we evaluated associations of germline variants with NSND YA OSCC in patients of European ancestry. We excluded cases of non-European ancestry, as linkage disequilibrium could have introduced false positives in our small cohort. Our results suggest that no single germline variant explains NSND YA OSCC occurrence and no variant with genome-wide significance ($p < 5.0 \times 10^{-8}$) was found.

Our main objective in this study was to identify potential rare variants associated with NSND YA OSCC which would have the potential to be causal factors. To define a variant as rare, we consider variants with a MAF < 0.01 in every genetic ancestry group reported in the All of Us cohort. As our study had only 10 NSND YA OSCC patients, we used a suggestive significance threshold of $p < 1.0 \times 10^{-5}$, as we were aware that we might lack the necessary statistical power to identify potential rare variants with a more stringent genome-wide significance threshold. Indeed, to attain genome-wide significance, a MAF of 0.5 (10/20 alleles) would be needed in the 10 patients NSND YA OSCC test group, for a variant with a MAF of 0.01 (3/300 alleles) in the control group. For a variant with no occurrence (0/300 alleles) in the control group, a MAF of 0.4 (8/20 alleles) would be necessary to attain genome-wide significance. This is hardly realistic and we chose to use a suggestive significance threshold for this exploratory study. We also tested the larger NSND OSCC and YA OSCC groups in which smaller effect sizes could be detected.

Only one of reported variants for NSND YA OSCC patients, rs191595756, has a reported MAF < 0.01 in every ancestry group. Interestingly, this variant is located in an intron of LINC00944 (Long Intergenic Non-Protein Coding RNA 944) and an exon of LINC02824 (Long Intergenic Non-Protein Coding RNA 2824). Multiple long non-coding RNAs have been shown to have oncogenic or tumor suppressive effects in OSCC, regulating multiple pathways such as cell cycle, angiogenesis and apoptosis [36]. LINC00944 has been studied in multiple cancers [37–39]. The LINC00944 gene has been

shown to actually encode a small 102-amino acid peptide, which represses transcriptional activity of MYC in vitro in melanoma cells [40]. The importance of MYC in head and neck squamous cell carcinoma is well established, with gene amplification occurring in approximately 12% of cases [41]. LINC00944 has also been shown to be upregulated in head and neck squamous cell carcinoma in comparison to normal adjacent tissues [42]. Overexpression of LINC00944 in the OSCC cell lines SCC-25 and ORL-48T increased cell migration and cell invasion in Transwell assays [42]. LINC00944 was also associated to pro-inflammatory signaling and is upregulated in EBV oral cancer cell lines [42]. The importance of EBV in head and neck cancer is primarily associated with a large increase in the risk of developing nasopharyngeal carcinoma, with EBV-associated OSCC being of rare occurrence [43, 44]. The association between MYC and EBV is well established. MYC regulates the switch between EBV latent and lytic stages [45], and also multiple pathways associated with immune evasion [46]. LINC02824 is a mostly unknown long non-coding RNA, being reported only to be positively associated with the number of regulatory T cells in systemic lupus erythematosus [47].

We also evaluated variants associated with all-ages NSND OSCC and with YA OSCC. No rare variant was found to be associated in both cases, even at the suggestive significance threshold, despite having more cases than in the NSND YA group (respectively 29 and 22, versus only 10 for NSND YA). Interestingly, while three of the five variants reported for the NSND YA group had a highest reported MAF < 0.02 , each of the eleven reported variants for the NSND and the YA group were common variants with a highest reported MAF over 0.05.

Our genome-wide association analysis results suggest that no single germline variant explains NSND YA OSCC occurrence and that a polygenic risk score could be more appropriate to evaluate NSND YA OSCC risk. A larger multicentric study could potentially identify a polygenic risk score predictive of NSND YA OSCC risk, as our cohort including only 10 NSND YA OSCC cases is not large enough for an analysis with NSND YA analyzed as a binary trait. However, calculating a PRS score for age at OSCC diagnosis as a numerical variable as an exploratory analysis appeared reasonable. It also allowed to evaluate whether genetic susceptibility to YA OSCC is similar in NSND and ESD OSCC in our cohort.

Our PRS score for age at OSCC diagnosis was shown to be independent of NSND or ESD status. This result is consistent with previous results showing that association analyses considering OSCC as a whole tend to identify variants associated with tobacco and alcohol consumption [22]. Therefore, the NSND YA OSCC

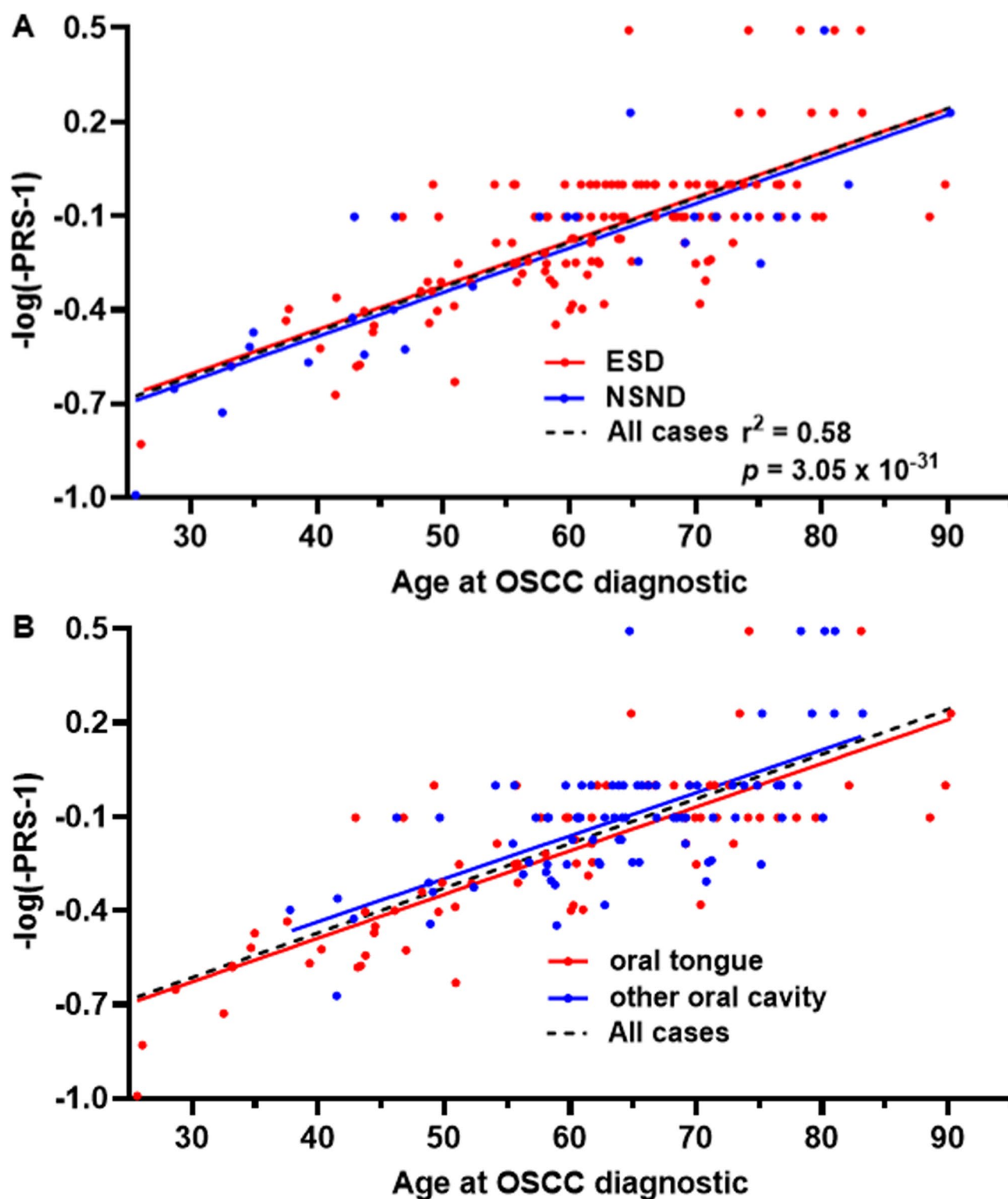


Fig. 2 Association between age at diagnosis of oral squamous cell cancer (OSCC) and transformed polygenic risk score (PRS). Association of the age at OSCC diagnosis with $-\log(-PRS-1)$ for **(A)** never-smoker non-drinker (NSND), ever-smoker and/or drinker (ESD) and all cases, and **(B)** oral tongue, other oral cavity and all cases. No significant difference is observed between the NSND and ESD slopes and intercepts. No significant difference is observed between the oral tongue and other oral cavity slopes. The displayed linear regression r^2 and p -value are calculated for all cases

group should be evaluated independently, and a larger cohort will be needed to perform this analysis, as our study sample size was too small to calculate a PRS specifically for the NSND YA OSCC group. YA OSCC predominantly involves oral tongue cancers [9], but association of PRS score with age at diagnosis was comparable in oral tongue and other oral cavity cases.

This study has multiple limitations. It was a retrospective, monocentric study with data extracted from clinical records and not from study questionnaires. The cohort includes exclusively patients of European ancestry, with the majority being French Canadians, which limits generalization and could introduce subtle population stratification in the study. Additionally, the total number of participants was limited for such a study, and no validation cohort was available for the PRS score. A sample size of 160 is small for PRS and threshold optimization increases the risk of overfitting. HPV infection status was not recorded, as it had not been tested for most patients. Previous studies report a low prevalence of HPV in OSCC, particularly in YA cases, but the lack of HPV status data could affect the interpretation of this study. Multi-allelic variants and variants located on the X and Y chromosomes were not evaluated in this study.

Conclusions

The main objective of this study was to evaluate whether rare germline variants are associated with NSND YA OSCC cases. One rare variant, rs191595756, was suggestively associated with NSND YA OSCC, but did not reach genome-wide significance. No variant with $p < 5.0 \times 10^{-8}$ was found. Our findings suggest that no single germline variant explains NSND YA OSCC occurrence. A larger multicentric study is needed to identify a polygenic risk score predictive of NSND YA OSCC risk.

Abbreviations

AJCC	American Joint Committee on Cancer
CHUM	Centre Hospitalier de l'Université de Montréal
CRCHUM	CHUM Research Center
ESD	Ever-smoker and/or drinker
GATK	Genome analysis toolkit
HLA	Human leukocyte antigen
HPV	Human papillomavirus
ICD-10	International Classification of Diseases, Tenth Revision
LINC00944	Long intergenic non-protein coding RNA 944
LINC02824	Long intergenic non-protein coding RNA 2824
MAF	Minor allele frequency
MCV	Merkel cell polyomavirus
MOHCCN	Marathon of Hope Cancer Centres Network
NSND	Never-smoking, non-drinking
OSCC	Oral squamous cell carcinoma
PRS	Polygenic risk score
VCF	Variant Call Format
WGS	Whole genome sequencing
YA	Young adults

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-026-16342-2>.

Supplementary Material 1. Table S1. Clinical characteristics of the included participants.

Supplementary Material 2. Table S2. Germline variants with $p < 1 \times 10^{-5}$ identified in the genome-wide association study comparing never-smoking non-drinking (NSND) young adults (YA) oral squamous cell carcinoma (OSCC) patients with other OSCC patients.

Supplementary Material 3. Table S3. Germline variants with $p < 1 \times 10^{-5}$ identified in the genome-wide association analysis comparing all-ages never-smoking, non-drinking (NSND) oral squamous cell carcinoma (OSCC) patients with ever-smoker and/or drinker (ESD) OSCC patients.

Supplementary Material 4. Table S4. Germline variants with $p < 1 \times 10^{-5}$ identified in the genome-wide association analysis comparing young adults (YA) oral squamous cell carcinoma (OSCC) patients with older OSCC patients.

Supplementary Material 5. Table S5. Germline variants included in the polygenic score for age at oral squamous cell carcinoma (OSCC) diagnosis.

Supplementary Material 6. Table S6 Univariate and multivariate analysis of overall survival for the study cohort.

Supplementary Material 7. Figure S1. Q-Q plots for the genome-wide association analysis of for never-smoking non-drinking (NSND) young adults (YA) oral squamous cell carcinoma (OSCC).

Supplementary Material 8. Figure S2. Linear regressions between age at OSCC diagnosis and PRS for NSND, ESD and all cases.

Supplementary Material 9. Figure S3. Evaluation of PRS association with pathologic TNM.

Acknowledgements

We thank the Christopoulos and Rodier laboratory members for valuable comments and discussions. We thank, Liliane Meunier and Veronique Barres of the Molecular Pathology core facility of the CRCHUM for performing DNA quality control analysis and DNA extraction preparation from paraffin blocs. We would like to acknowledge Pascale Marquis and her colleagues at McGill University's Canadian Centre for Computational Genomics (C3G) for their work in performing secondary bioinformatics analyses and quality control of the whole-genome sequencing data used in this study. We would also like to acknowledge Elena Fraiji support for data deposition. This research, as well as the biobanking of biological material and data, was made possible through a collaboration with the Réseau de recherche sur le cancer (RRCancer), which is affiliated with the Canadian Tumor Repository Network (CTRNet). The results published here are in based upon data generated by the Terry Fox Research Institute Marathon of Hope Cancer Centres Network and the All of Us data browser. We gratefully acknowledge the Terry Fox Research Institute Marathon of Hope Cancer Centres Network and All of Us participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's All of Us Research Program for making available the participant data examined in this study.

Authors' contributions

GBC and AC conceived and designed the study. GBC, APV, MB, FK, TA, EB, LG, PT, OEG, EF, PFNT, AD, HB and AC assisted in or provided support for recruitment of participants and sample collection. GBC, APV, MB and FK performed sample processing and DNA extractions. GBC performed statistical analysis. GBC, APV, MB, FK, APC, FR, HB, and AC analyzed and interpreted the data. GBC, APV, MB, FK, TA, EB, LG, PT, OEG, EF, PFNT, AD, APC, FR, HB, and AC critically reviewed the manuscript for valuable intellectual content. HB and AC obtained funding. FR, HB and AC supervised the study. GBC wrote the first draft of the manuscript. All authors read and approved the final manuscript prior to submission.

Funding

This research was funded by the Institut du cancer de Montréal, by the Dr Azar-Angelil Research Chair in Head and Neck Oncology of the Université de Montréal and by the Cancer Research Society (1274287). The data presented in this paper were partially funded by the Terry Fox Research Institute's Marathon of Hope Cancer Centres Network.

Data availability

The datasets generated and/or analysed during the current study are available in the European Genome-Phenome Archive (EGAS50000001809).

Declarations

Ethics approval and consent to participate

The Marathon of Hope project and the institutional head and neck cancer biobank were approved by the *Centre de recherche du Centre hospitalier de l'Université de Montréal* research ethics board (21.137 and 09.254). All participants provided written informed consent. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM), 1051 rue Sanguinet, Montreal, QC H2X 3E4, Canada

²Institut du cancer de Montréal, Montreal, QC, Canada

³Otolaryngology-Head and Neck Surgery Service, Department of Surgery, Centre hospitalier de l'Université de Montréal (CHUM), Montreal, QC, Canada

⁴Department of Pathology, CHUM, Montreal, QC, Canada

⁵Department of Radiation Oncology, CHUM, Montreal, QC, Canada

⁶Hematology-Oncology Service, Department of Medicine, CHUM, Montreal, Canada

⁷École de Technologie Supérieure, Montreal, QC, Canada

⁸Département de Radiologie, Radiation Oncology and Nuclear Medicine, Université de Montréal, Montréal, Canada

Received: 12 November 2025 / Accepted: 8 June 2026

Published online: 12 June 2026

References

1. Amsterdam JT, Strawitz JG. Squamous cell carcinoma of the oral cavity in young adults. *J Surg Oncol*. 1982;19(2):65–8.
2. Shemen LJ, Klotz J, Schottenfeld D, Strong EW. Increase of tongue cancer in young men. *JAMA*. 1984;252(14):1857.
3. Macfarlane GJ, Boyle P, Scully C. Oral cancer in Scotland: changing incidence and mortality. *BMJ*. 1992;305(6862):1121–3.
4. Schantz SP, Yu GP. Head and neck cancer incidence trends in young Americans, 1973–1997, with a special analysis for tongue cancer. *Arch Otolaryngol Head Neck Surg*. 2002;128(3):268–74.
5. Annertz K, Anderson H, Björklund A, Moller T, Kantola S, Mork J, et al. Incidence and survival of squamous cell carcinoma of the tongue in Scandinavia, with special reference to young adults. *Int J Cancer*. 2002;101(1):95–9.
6. Shiboski CH, Schmidt BL, Jordan RC. Tongue and tonsil carcinoma: increasing trends in the U.S. population ages 20–44 years. *Cancer*. 2005;103(9):1843–9.
7. Patel SC, Carpenter WR, Tyree S, Couch ME, Weissler M, Hackman T, et al. Increasing incidence of oral tongue squamous cell carcinoma in young white women, age 18 to 44 years. *J Clin Oncol*. 2011;29(11):1488–94.
8. Silverman S. Jr. Demographics and occurrence of oral and pharyngeal cancers. The outcomes, the trends, the challenge. *J Am Dent Assoc*. 2001;132(Suppl):S7–11.
9. Dahlstrom KR, Little JA, Zafereo ME, Lung M, Wei Q, Sturgis EM. Squamous cell carcinoma of the head and neck in never smoker-never drinkers: a descriptive epidemiologic study. *Head Neck*. 2008;30(1):75–84.
10. Toporcov TN, Znaor A, Zhang ZF, Yu GP, Winn DM, Wei Q, et al. Risk factors for head and neck cancer in young adults: a pooled analysis in the INHANCE consortium. *Int J Epidemiol*. 2015;44(1):169–85.
11. Li R, Faden DL, Fakhry C, Langelier C, Jiao Y, Wang Y, et al. Clinical, genomic, and metagenomic characterization of oral tongue squamous cell carcinoma in patients who do not smoke. *Head Neck*. 2015;37(11):1642–9.
12. Salem A. Dismissing links between HPV and aggressive tongue cancer in young patients. *Ann Oncol*. 2010;21(1):13–7.
13. Lingen MW, Xiao W, Schmitt A, Jiang B, Pickard R, Kreinbrink P, et al. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. *Oral Oncol*. 2013;49(1):1–8.
14. Muller S, Tilakaratne WM. Update from the 5th Edition of the World Health Organization Classification of Head and Neck Tumors: Tumours of the Oral Cavity and Mobile Tongue. *Head Neck Pathol*. 2022;16(1):54–62.
15. Kabeya M, Furuta R, Kawabata K, Takahashi S, Ishikawa Y. Prevalence of human papillomavirus in mobile tongue cancer with particular reference to young patients. *Cancer Sci*. 2012;103(2):161–8.
16. Poling JS, Ma XJ, Bui S, Luo Y, Li R, Koch WM, et al. Human papillomavirus (HPV) status of non-tobacco related squamous cell carcinomas of the lateral tongue. *Oral Oncol*. 2014;50(4):306–10.
17. Mulder FJ, Klufah F, Janssen FME, Farshadpour F, Willems SM, de Bree R, et al. Presence of Human Papillomavirus and Epstein-Barr Virus, but Absence of Merkel Cell Polyomavirus, in Head and Neck Cancer of Non-Smokers and Non-Drinkers. *Front Oncol*. 2020;10:560434.
18. Garavello W, Foschi R, Talamini R, La Vecchia C, Rossi M, Dal Maso L, et al. Family history and the risk of oral and pharyngeal cancer. *Int J Cancer*. 2008;122(8):1827–31.
19. Pickering CR, Zhang J, Neskey DM, Zhao M, Jasser SA, Wang J, et al. Squamous cell carcinoma of the oral tongue in young non-smokers is genomically similar to tumors in older smokers. *Clin Cancer Res*. 2014;20(14):3842–8.
20. Jiang R, Gao MZ, Chen M, Weatherspoon DJ, Watts TL, Osazuwa-Peters N. Genetic and Molecular Differences in Head and Neck Cancer Based on Smoking History. *JAMA Otolaryngol Head Neck Surg*. 2025;151(4):379–88.
21. Lesseur C, Diergaarde B, Olshan AF, Wunsch-Filho V, Ness AR, Liu G, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet*. 2016;48(12):1544–50.
22. Ebrahimi E, Sangphukieo A, Park HA, Gaborieau V, Ferreiro-Iglesias A, Diergaarde B, et al. Cross-ancestral GWAS identifies 29 novel variants across Head and Neck Cancer subsites. *medRxiv*. 2024.
23. Nikkila R, Makitie A, Joensuu H, Markkanen S, Elenius K, Monni O, et al. Novel Genetic Risk Variants Associated with Oral Tongue Squamous Cell Carcinoma. *Head Neck Pathol*. 2025;19(1):45.
24. Bau DT, Liu TY, Tsai CW, Chang WS, Gu J, Yang JS, et al. A Genome-Wide Association Study Identified Novel Genetic Susceptibility Loci for Oral Cancer in Taiwan. *Int J Mol Sci*. 2023;24(3).
25. Terry Fox Research Institute Marathon of Hope Cancer Centers Network. Electronic address mbc, Terry Fox Research Institute Marathon of Hope Cancer Centers N. The Terry Fox Research Institute Marathon of Hope Cancer Centres Network: A pan-Canadian precision oncology initiative. *Cancer Cell*. 2025;43(4):587–92.
26. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013;1303.3997v1 [q-bio.GN].
27. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568–76.
28. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016;44(11):e108.
29. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591–4.
30. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2).
31. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
33. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*. 2023;39(1).

34. All of Us Research, Program I, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The All of Us Research Program. *N Engl J Med*. 2019;381(7):668–76.
35. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*. 2020;15(9):2759–72.
36. Tang J, Fang X, Chen J, Zhang H, Tang Z. Long Non-Coding RNA (lncRNA) in Oral Squamous Cell Carcinoma: Biological Function and Clinical Application. *Cancers (Basel)*. 2021;13(23).
37. Li Y, Jiang T, Zhou W, Li J, Li X, Wang Q, et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun*. 2020;11(1):1000.
38. Chen C, Zheng H. lncRNA LINC00944 Promotes Tumorigenesis but Suppresses Akt Phosphorylation in Renal Cell Carcinoma. *Front Mol Biosci*. 2021;8:697962.
39. de Santiago PR, Blanco A, Morales F, Marcelain K, Harismendy O, Sjöberg Herrera M, et al. Immune-related lncRNA LINC00944 responds to variations in ADAR1 levels and it is associated with breast cancer prognosis. *Life Sci*. 2021;268:118956.
40. Ren Z, Liu J, Chang X, Yang X, Zhang Y, Zhang X, et al. A peptide encoded by LINC00944 suppresses the growth of melanoma cells by diminishing EP400-MYC interaction. *Biochem Pharmacol*. 2025;231:116652.
41. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517(7536):576–82.
42. Srisathaporn S, Ekalaksananan T, Heawchaiyaphum C, Aromseree S, Maranon DG, Altina NH et al. EBV-Induced LINC00944: A Driver of Oral Cancer Progression and Influencer of Macrophage Differentiation. *Cancers (Basel)*. 2025;17(3).
43. Ji MF, He YQ, Tang MZ, Xue WQ, Yu X, Diao H, et al. Epstein Barr virus antibody and cancer risk in two prospective cohorts in Southern China. *Nat Commun*. 2025;16(1):5940.
44. Guidry JT, Birdwell CE, Scott RS. Epstein-Barr virus in the pathogenesis of oral cancers. *Oral Dis*. 2018;24(4):497–508.
45. Guo R, Jiang C, Zhang Y, Govande A, Trudeau SJ, Chen F, et al. MYC Controls the Epstein-Barr Virus Lytic Switch. *Mol Cell*. 2020;78(4):653–69. e8.
46. Mahdavi P, Panahipour Javaherdehi A, Khanjanpoor P, Aminian H, Zakeri M, Zafarani A, et al. The role of c-Myc in Epstein-Barr virus-associated cancers: Mechanistic insights and therapeutic implications. *Microb Pathog*. 2024;197:107025.
47. Bu YJ, Cen X, Wang YQ, Fan R, Zhang F, Liu YQ, et al. Study on the expression changes of lncRNA in patients with systemic lupus erythematosus and its correlation with Treg cells. *Clin Rheumatol*. 2024;43(3):993–1002.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.